09/806326

SUSCEPTIBILITY LOCUS FOR OSTEOARTHRITIS

Technical Field

This invention relates to the identification of chromosomal regions linked to susceptibility to osteoarthritis using linkage and association analysis.

Background of Invention

Osteoarthritis (OA) is a debilitating disease involving degeneration of the articular cartilage of synovial joints[5,6]. Although OA has long been considered an inevitable consequence of ageing, it has become increasingly apparent that OA does have a genetic component. Early-onset forms of the disease are associated with several osteochondrodysplasias, rare diseases involving abnormal bone and cartilage development that are transmitted as Mendelian traits[7]. However, the OA in these conditions is secondary to the main dysplastic phenotype. The common late-onset form of the disease (idiopathic OA) often has no obvious environmental (i.e. injury) or characteristic physical cause and does not demonstrate a clear mode of inheritance.

Over the last 10 years, many genes for single gene or monogenic diseases, which are relatively rare in the population, have been positioned by linkage analysis in families, and localised to a small enough region to allow identification of the gene. The latter sublocalisation and fine mapping can be carried out in single gene rare diseases because recombinations within families define the boundaries of the minimal interval beyond any doubt. In contrast, in common diseases such as osteoarthritis, diabetes or asthma the presence of the disease mutation does not always coincide with the development of the disease: disease susceptibility mutations in common disorders provide risk of developing of the disease, and this risk is usually much less than 100%.

Hence, susceptibility genes in common diseases cannot be localised using recombination events within families, unless tens of thousands of families are available to fine map the locus. Because collections of this size are impractical, investigators are contemplating the use of association mapping, which relies on historical recombination events during the history of the population from which the families came from.

Association mapping has been used in over a dozen examples of rare single gene traits, and particularly in genetically isolated populations such as Finland to fine map disease mutations. Nevertheless, association mapping is fundamentally different from straightforward linkage mapping because even though the degree of association between two markers or a marker and a disease mutation is proportional to the physical distance along the chromosome this relationship can be unpredictable because it is dependent on the allele frequencies of the markers, the history of the population and the age and number of mutations at the disease locus. For rare, highly penetrant single gene diseases there is usually one major founder chromosome in the population under study, making it relatively feasible to locate an interval that is smaller than one that can be defined by standard recombination events within living families. The resolution of this method in monogenic diseases in which there is one main founder chromosome is certainly less than 2cM, and in certain examples the resolution is down to 100 kb of DNA (Hastbacka *et al*. (1994) Cell 78,1-20).

In common diseases like OA, which are caused by a number of genes or polygenes acting together in concert the population frequency of the disease allele may be very high, perhaps exceeding 50%, and there are likely to be several founder chromosomes, all of which impart risk, and not a 100%

certainty of disease development. Because association mapping is dependent on unpredictable parameters, and because founder chromosomes will be several and common in frequency in the general population, the task of fine mapping polygenes is currently one of some controversy, and many doubt the feasibility at all of a systematic genetic approach using a combination of linkage and association mapping. Recently, Risch and Marakandis have provided some mathematical background to the feasibility of association mapping in complex diseases (*Science* 273 1516-1517, 1996) but they did not take into account the effect of multiple founder chromosomes.

It has often been noted in epidemiological studies that there is a female preponderance for OA[5,6]. This may be accounted for by differential effects on the two sexes of environmental factors. However, a Finnish twin study has suggested that genetic susceptibility may be greater in women than men[9]. This result has recently been supported by a segregation analysis[3]. Not only have differences in risk between females and males been reported but it has also been suggested that there are differences in heritability between joint groups[2,10,11]. These differences could be the result of genetic locus heterogeneity between the different joints.

Summary of the Invention

The present inventors have identified regions on chromosome 2q that may harbour susceptibility loci for OA. This region was identified following a two-stage non-parametric linkage analysis. The first stage involved a random screen of the genome using 272 microsatellite markers in 297 OA families. The second stage was more selective and involved genotyping an additional 184 families for those markers that demonstrated moderate evidence of linkage in stage 1. This revealed one microsatellite on chromosome 2 for which the

evidence for linkage increased as the number of families studied increased. Finer mapping in and around this microsatellite was performed which provided enhanced evidence for linkage and enabled linked regions to be defined. Stratification analysis suggested that the chromosome 2 loci may have differential penetrance between males and females and between the two different joint groups examined (hips and knees).

Linkage on chromosome 2 may encompass three distinct loci: a locus close to D2S114, a locus in-between D2S2330 and D2S326 and a locus close to or in-between D2S117 and D2S325. None of the single or multipoint LOD scores for these three regions are particularly high, with the highest single point LOD score being 2.36 at marker D2S202 and the highest MMLS (Maximum multipoint log score) being 2.07 in-between markers D2S117 and D2S72, both in hip-only pairs. What is more striking about our linkage results for chromosome 2 is that there is overlap with regions that other groups have previously identified as potentially harbouring OA susceptibility loci[12,13]. This provides corroboratory evidence that the linkages to chromosome 2 that we have detected are real.

Some evidence has also been obtained for linkage with the polymorphic chromosome 6 marker D6S273 and the polymorphic X chromosome marker DXS1068. This may indicate the presence of loci on these chromosomes that also have a role in OA susceptibility. These loci may interact with the loci identified on chromosome 2 and play a role in the differential penetration observed between males and females.

Open reading frames (ORFs) are stretches of genetic sequence which are candidates for being expressed genes. They can be identified by the presence of sequence elements which are

characteristic of coding sequence, such as sequence elements from exon-intron boundaries, transcriptional initiation and termination motifs and start and stop codons. Because large amounts of sequence can be screened rapidly for these elements, the identification of ORFs is commonly an initial step in the discovery of novel genes.

Microsatellite marker loci are designated in this application according to the nomenclature conventional in the field of Human Genetics. This provides a unique designation which specifically and unambiguously identifies each marker locus. Mapping data for any particular marker locus is readily available from conventional sources, such as the Gopher server at the Human Genome Mapping Project Resource Centre (Host = gopher.hgmp.mrc.ac.uk; Port = 70 or URL: gopher://gopher.hgmp.mrc.ac.us:70/ or by anonymous ftp from ftp.hgmp.mrc.uk:/Oxford_Primers).

The present invention relates to chromosomal regions linked to genetic sequences which affect susceptibilty to osteoarthritis.

A first aspect of the present invention is a method for identifying individuals susceptible to osteoarthritis comprising obtaining a sample of genomic DNA and detecting the presence or absence of any one of the 190bp and 200bp alleles of D2S325 from chromosome 2.

Another aspect of the present invention is a method for identifying male individuals susceptible to osteoarthritis comprising obtaining a sample of genomic DNA and detecting the presence or absence of any one of the 190bp and 200bp alleles of D2S325 from chromosome 2.

Another aspect of the present invention is a method for identifying male individuals susceptible to osteoarthritis comprising obtaining a sample of genomic DNA and detecting the presence or absence of the 208bp allele of D2S117 from chromosome 2.

A further aspect of the present invention is a method for identifying female individuals susceptible to osteoarthritis comprising obtaining a sample of genomic DNA and detecting the presence or absence of any one of the 192bp, 202bp and 208bp alleles of D2S117 from chromosome 2.

A further aspect of the present invention is a method for identifying male individuals susceptible to osteoarthritis of the hip comprising obtaining a sample of genomic DNA and detecting the presence or absence of the 208 bp allele of D2S117 from chromosome 2.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising screening genomic libraries with sequence from any one of the 190bp and 200bp alleles of D2S325 and identifying open reading frames in regions adjacent to said allele.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising screening a genomic library from an individual who is homozygote for any one of the 190bp and 200bp alleles of D2S325 and identifying open reading frames in regions adjacent to said allele.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising screening a genomic library from an individual who is homozygote for any one of the 190bp and the 200bp alleles

of D2S325 and identifying open reading frames located within 500 Kb of D2S325, or more preferably within 100 Kb of D2S325, or even more preferably within 50 Kb of D2S325 or most preferably within 10 Kb of D2S325.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising identifying open reading frames in regions adjacent to D2S325 and comparing said open reading frames in individuals carrying any one of the 190 bp and 200bp alleles of D2S325 with said open reading frames in individuals with other alleles of D2S325.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising identifying open reading frames located within 500 Kb of D2S325, or more preferably within 100 Kb of D2S325, or even more preferably within 50 Kb of D2S325 or most preferably within 10 Kb of D2S325 and comparing said open reading frames in individuals carrying any one of the 190 bp and 200 bp alleles of D2S325 with said open reading frames in individuals with other alleles of D2S325.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising screening genomic libraries with sequence from the 192bp and 202bp alleles of D2S117 and identifying open reading frames in regions adjacent to said allele.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising screening a genomic library from an individual who is homozygote for any one of the 192bp, the 202bp and the 208bp alleles of D2S117 and identifying open reading frames in regions adjacent to said allele

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising screening a genomic library from an individual who is homozygote for any one of the 192bp, 202bp and 208bp alleles of D2S117 and identifying open reading frames located within 500 Kb of D2S117, or more preferably within 100 Kb of D2S117, or even more preferably within 50 Kb of D2S117 or most preferably within 10 Kb of D2S117.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising identifying open reading frames in regions adjacent to D2S117 and comparing said open reading frames in individuals carrying any one of the 192bp, 202bp and 208bp alleles of D2S117 with said open reading frames in individuals with other alleles of D2S117.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising identifying open reading frames located within 500 Kb of D2S117, or more preferably within 100 Kb of D2S117, or even more preferably within 50 Kb of D2S117 or most preferably within 10 Kb of D2S117 and comparing said open reading frames in individuals carrying any one of the 192bp, 202bp and 208bp alleles of D2S117 with said open reading frames in individuals with other alleles of D2S117.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising screening genomic libraries with sequence from any one of the 190bp and 200bp alleles of D2S325 and identifying open reading frames located within 500 Kb of D2S325, or more preferably within 100 Kb of D2S325, or even more preferably within 50 Kb of D2S325 or most preferably within 10 Kb of D2S325.

Another aspect of the present invention is a method for isolating genetic loci associated with susceptability to OA comprising screening genomic libraries with sequence from any one of the 192bp, 202bp and 208bp alleles of D2S117 and identifying open reading frames located within 500 Kb of D2S117, or more preferably within 100 Kb of D2S117, or even more preferably within 50 Kb of D2S117 or most preferably within 10 Kb of D2S117.

Another aspect of the present invention is the use of the 190bp and 200bp alleles of D2S325 as markers for the identification of loci influencing susceptibility to OA.

Another aspect of the present invention is the use of any one of the 192bp, 202bp and 208bp alleles of D2S117 as a marker for the identification of loci influencing susceptibility to OA.

Another aspect of the present invention is a method for mapping loci which affect susceptability to OA by comparing genomic regions containing the 208 bp allele of D2S117 with genomic regions containing other alleles of D2S117.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at least two of which have clinical symptoms of osteoarthritis. analysing a region of their genomic DNA comprising a polymorphic marker, said region being located on chromosome 2q between D2S117 and D2S325, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at least two of which have clinical symptoms of osteoarthritis. analysing a region of their genomic DNA comprising a polymorphic marker, said region being located on chromosome 2q between D2S202 and D2S72, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at least two of which have clinical symptoms of osteoarthritis analysing a region of their genomic DNA comprising a polymorphic marker, said region being located on chromosome 2q between D2S117 and D2S325, and additionally analysing one or more of the following; a genomic region comprising the polymorphic marker D6S273 and a genomic region comprising the polymorphic marker DXS1068, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at least two of which have clinical symptoms of osteoarthritis analysing a region of their genomic DNA comprising a polymorphic marker, said region being located on chromosome 2q between D2S202 and D2S72, and additionally analysing one or more of the following; a genomic region comprising the

polymorphic marker D6S273 and a genomic region comprising the polymorphic marker DXS1068, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at least two of which have clinical symptoms of osteoarthritis. analysing a region of their genomic DNA comprising the polymorphic marker D2S114, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at least two of which have clinical symptoms of osteoarthritis. analysing a region of their genomic DNA comprising the polymorphic marker D2S114, and additionally analysing one or more of the following; a genomic region comprising the polymorphic marker D6S273 and a genomic region comprising the polymorphic marker DXS1068, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at

least two of which have clinical symptoms of osteoarthritis. analysing a region of their genomic DNA comprising a polymorphic marker, said region being located on chromosome 2q between D2S330 and D2S326, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at least two of which have clinical symptoms of osteoarthritis. analysing a region of their genomic DNA comprising a polymorphic marker, said region being located on chromosome 2q between D2S330 and D2S326,and additionally analysing one or more of the following; a genomic region comprising the polymorphic marker D6S273 and a genomic region comprising the polymorphic marker DXS1068, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at least two of which have clinical symptoms of osteoarthritis. analysing a region of their genomic DNA comprising any one of the polymorphic markers; D2S202, D2S72, D3S1266, D4S231, D4S415, D6S260, D6S273, D6S286, D6S281, D7S669, D7S530, D11S907, D11S903, D11S901, D17S807, D17S789, DXS1068, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual

susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for determining individual susceptibility to osteoarthritis comprising obtaining sample genomic DNA from siblings, at least two of which have clinical symptoms of osteoarthritis. analysing a region of their genomic DNA comprising any one of the polymorphic markers D6S273, DXS1068, identifying allele sharing between the siblings as defined by a maximum log of the odds (LOD) score of greater than 1 and a p-value of less then 0.25, and determining individual susceptibility to osteoarthritis by reference to the allele sharing.

Another aspect of the present invention is a method for identifying loci conferring susceptibilty to osteoarthritis comprising screening a genomic library with genetic sequence derived from one or more of the following polymorphic markers; D2S202, D3S1266, D4S231, D4S415, D6S260, D6S273, D6S286, D6S281, D7S669, D7S530, D11S907, D11S903, D11S901, D17S807, D17S789, DXS1068 and identifying those isolated clones which additionally contain open reading frames

Another aspect of the present invention is a method for identifying loci conferring susceptibilty to osteoarthritis comprising screening a genomic library with genetic sequence derived from one or more of the following polymorphic markers; D2S202, D3S1266, D4S231, D4S415, D6S260, D6S273, D6S286, D6S281, D7S669, D7S530, D11S907, D11S903, D11S901, D17S807, D17S789, DXS1068 and identifying open reading frames located within 500 Kb, more preferably within 100 Kb, even more preferably within 50 Kb or most preferably within 10 Kb of any of these polymorphic markers.

Another aspect of the present invention is a method for

identifying loci conferring susceptibilty to osteoarthritis comprising screening a genomic library with genetic sequence derived from one or more of the polymorphic markers D6S273, DXS1068 and identifying open reading frames located within 500 Kb, more preferably within 100 Kb, even more preferably within 50 Kb or most preferably within 10 Kb of either of these polymorphic markers.

Description of Drawings

Figure 1 shows a multipoint log of the odds (LOD) analysis on the unstratified data.

Figure 2 shows a multipoint log of the odds (LOD) analysis on stratified data from male sibling pairs only.

Figure 3 shows a multipoint log of the odds (LOD) analysis on stratified data from hip pairs only.

Figure 4 shows a multipoint log of the odds (LOD) analysis on stratified data from male/hip pairs only.

Detailed Description of Invention

Linkage Analysis

The present inventors collected 481 families in which at least two siblings have undergone joint replacement surgery of the hip or knee for severe, end-stage idiopathic OA (Table 1). Due to the late onset of the disease, none of these families contained parents who could participate in the study. In stage 1, 272 microsatellite markers were genotyped in 297 of the 481 families. The microsatellites were essentially those used by Reed *et al*[8] with the replacement of certain markers that amplified unreliably. Sixteen markers from stage 1 had evidence of linkage at $p \leq 0.05$ (Table 2). These markers were then genotyped in the remaining 184 families. None of the 16 markers had a $p \leq 0.05$ in this second stage although three had a $p \leq 0.10$: D2S202 ($p=0.07$), D11S903 ($p=0.07$) and D11S901 ($p=0.10$) (Table 2). When the data for stages 1 and 2 were combined and compared to stage 1 only,

the combined p-value decreased for 4 of the 16 markers: D2S202 (p=0.009 for combined vs. 0.036 for stage 1), D11S907 (p=0.019 for combined vs. 0.05 for stage 1), D11S903 (p=0.004 for combined vs. 0.017 for stage 1) and D11S901 (p=0.0003 for combined vs. 0.0004 for stage 1) increasing the number of families had therefore increased the evidence for linkage at these 4 markers. The present invention is related to the chromosome 2 marker (D2S202) thus identified.

Although D2S202 was the only marker on chromosome 2 that had a $p \leq 0.05$ in stage 1, one other chromosome 2 marker had a $p \leq 0.10$: D2S72 (p=0.07). This marker is approximately 2cM from D2S202. When D2S72 was genotyped in stage 2 there was further evidence for linkage with a p-value of 0.024. When screens 1 and 2 were combined a p-value of 0.007 (LOD=1.26) was obtained for D2S72. This was a lower p-value than for D2S202 (0.007 vs. 0.009). The factor that two adjacent markers gave evidence of linkage at $p<0.01$ prompted us to genotype the stage 2 families for 6 other chromosome 2 markers from our original marker set that flanked D2S202 and D2S72. In addition, we genotyped 5 new markers for all 481 families. Table 3 lists the results for all 13 chromosome 2 markers that were genotyped in all 481 families. A multipoint analysis of this data gave a MMLS of 1.09 in-between markers D2S202 and D2S325 (Figure 1). Intriguingly, the multipoint analysis appears to highlight two additional distinct linkage regions, one located close to D2S114 (MMLS=0.95) and one located in-between D2S2330 and D2S326 (MMLS=0.91). Overall, these three peaks cover a distance greater than 60cM and therefore probably do not represent a single locus.

The results were stratified into six categories: those families that were affected female-only pairs (196 families), affected male-only pairs (male and/or female) (54 families),

affected female-only pairs who had undergone hip replacement but not knee replacement (female/hip pairs) (132 families) and affected male-only pairs who had undergone hip replacement but not knee replacement (male/hip pairs) (72 families) (Table 1). We did not stratify for female/knee pairs or male/knee pairs as the number of families were too low (21 and 8 respectively) and thus any significant results may simply be the result of stochastic variation.

Stratification of chromosome 2 revealed that the detected linkage was apparently accounted for by affected male-only pairs, with a single point LOD score of 1.79 (p=0.0021) at D2S202 and 1.70 (p=0.0025) at D2S72 (Table 4). The MMLS for male-only pairs was 1.02 in-between markers D2S117 and D2S72 (Figure 2). For affected female-only pairs there was no evidence of linkage at D2S202 (p=0.40), D2S72 (p=0.22) or any of the chromosome 2 markers except for marker D2S326 (p=0.019, 20cM proximal to D2S202). The linkage to chromosome 2 was restricted to pairs with hip-only disease with a p-value of 0.005 at D2S202 in hip-only pairs versus 0.34 in knee-only pairs. Again, however, the absence of linkage in knee-only pairs could be due to their relatively low number. The MMLS in hip-only pairs was 2.07, also in-between markers D2S117 and D2S72 (Figure 3). This linkage in hip-only pairs was restricted to males, with a p-value at D2S202 of 0.0016 in male/hips versus 0.20 in female/hips. The MMLS in male/hips was 1.54, also in-between markers D2S117 and D2S72 (Figure 4). Overall, these results suggest that the major linkage to chromosome 2 in our families is restricted to males with OA.

What is very striking about our multipoint analysis of chromosome 2 is that in the un-stratified and the stratified analysis there was at least one additional peak besides the major peak (Figures 1 to 4). These peaks may be false positives. Alternatively, they may be indicators of

additional independent loci. Our interpretation of these peaks can be enhanced by comparison to previous linkage reports of OA. Wright *et al*[12] have reported a linkage analysis of 12 markers on chromosome 2q in 66 nodal OA sib-pairs. Their 12 markers encompass over 65cM of 2q and in their study three were linked at p≤0.05: GCG, D2S326 and D2S126. These three markers flank D2S202, the marker that has the lowest p-value in our families: GCG and D2S326 are at least 20cM proximal to D2S202 whilst D2S126 is approximately 25cM distal to D2S202. Five of the 12 Wright *et al* markers are located in the interval in-between GCG/D2S326 and D2S126 and are therefore closer to D2S202. However, none of these supported linkage at p≤0.05. Of the 12 markers used by Wright *et al* D2S326 was the only one that was also used in our study. In our un-stratified families this marker had a LOD score of 0.45 (p=0.07) (Table 3). The majority of the families used by Wright *et al* contained female-only affected sib pairs. When we stratified for sex, the LOD score at D2S326 increased in our female-only pairs to 0.94 (p=0.019) but decreased in our male-only pairs to 0.18 (p=0.18) (Table 4). In our female/hips the LOD score was 0.74 (p=0.032). This female-specific linkage may therefore account for the peak observed in our un-stratified families that exists in-between markers D2S2330 and D2S326 (Figure 1). Intriguingly, this peak is also present when we stratify for hips-only (Figure 3) but is much less pronounced when we stratify for males-only (Figure 2) or male/hips (Figure 4). Our results for D2S326 point to a predominantly female-specific locus that is proximal to the predominantly male-specific locus that we have detected close to D2S202/D2S72.

Leppavuon *et al*[13] have also reported linkage to chromosome 2q in 42 affected sib pairs who have distal interphalangeal (DIP) OA, with a maximum LOD score of 4.06. They do not report the markers that they used in their study but instead give a cytogenetic location of 2q12-q14.2. This

is at least 70cM proximal to our major linkage at D2S202/D2S72, which cytogenetically map to 2q31. One of our chromosome 2 markers does map to 2q12-q14.2: D2S160. In our un-stratified families this marker had a LOD score of 0.65 (p=0.041) (Table 3). In our un-stratified data there is moderate evidence of linkage in-between markers D2S160 and D2S142, with a MMLS of 0.95 (Figure 1). This linkage extends from 2q14.1-2qq. Leppavuori *et al* did not report the sex-structure of their families. When we stratified our families for female-only and male-only affected pairs, neither group supported linkage at D2S160 (p=0.32 for female-only pairs and p=0.35 for male-only pairs) or at D2S114(p=0.29 for female-only pairs and p=0.11 for male-only pairs). However, in our hip-only pairs there was evidence of linkage at D2S160 (p=0.028) and D2S114 (p=0.022). Overall, it may be that our results do represent a confirmation of the linkage result of Leppavuori *et al* but that this locus confers only a moderate degree of susceptibility in our families. When we analyse our data as a whole (481 families) we have the power to detect linkage, as we do when we stratify for hip-only families (311 families). However, when we stratify for female-only (196 families) or male-only (102 families) we no longer have the power to detect linkage.

In conclusion our linkage analysis of chromosome 2, together with results previously reported, suggest that there may be up to three distinct OA susceptibility loci on 2q, one or more of which may demonstrate differential penetrance between males and females: a locus close to D2S114 that may be a susceptibility locus for females and males, a locus in-between D2S2330 and D2S326 that may be predominantly be a susceptibility locus for females and a locus close to or in-between D2S117 and D2S325 that may predominantly be a susceptibility locus for males.

<u>Markers from other chromosomes</u>

Since stratification analysis of chromosome 2 revealed apparent significant differences between the different categories examined, we also stratified 12 other markers that had a $p \leq 0.05$ in stage 1 and which were not on chromosome 2. The majority of these 12 markers will of course represent false positives from our stage 1 analysis: the p-value for each increased when the combined analysis was compared to stage 1 (Table 2). Nevertheless, they cannot all be discounted. Table 5 lists the stratification results for these 12 markers.

Two of the 12 markers are of interest when one considers the results of previous studies: D6S273 and DXS1068.

D6S273 maps close to the HLA complex on chromosome 6p and Pattrick *et al*[14] have reported association of nodal OA with a specific allele of the HLA-A gene. Furthermore, the COL11A2 gene, which encodes the α2 chain of the cartilage collagen type XI, is tightly linked to the HLA complex. This gene has been linked to and found to harbour causal mutations for the osteochondrodysplasia Stickler syndrome[15]. This syndrome has severe, precocious OA as one of its many phenotypic components.

Leppavuori *et al*[13] have reported linkage of DIP OA to chromosome 2q12-q14.2. A second linkage that they also reported maps to cytogenetic band Xp11.3. DXS1068 also maps to Xp11.3.

<u>Association Analysis</u>

Having established evidence for linkage to chromosome 2, the chromosome 2 markers were then tested for association. We used the Transmit software program[16] to test for association with linkage disequilibrium by the transmission disequilibrium test.

Since there are three possible distinct susceptibility loci on chromosome 2q we tested for disequilibrium for all 13 of our 2q markers. We tested the genotyping data without stratification and with the following stratification criteria: female-only pairs, male-only pairs, hip-only pairs, female/hip pairs and male/hip pairs.

Only D2S2330 demonstrated disequilibrium in the un-stratified data set, with a global $\chi^2_{5df} \equiv 11.43$ ($0.05>p>0.02$) (Table 6). This was accounted for by the significant distorted transmission of two alleles ($0.05>p>0.02$), one of which demonstrated excess transmission whilst the second demonstrated reduced transmission. In the stratified data disequilibrium was also positive for D2S2330 in the male/hips with global $\chi^2_{4df} \equiv 12.14$ ($0.02>p>0.01$). However, no one allele demonstrated significant distorted transmission and the allele that demonstrated excess transmission in the un-stratified data demonstrated reduced transmission in the male/hips. It is likely therefore that the positive disequilibrium results for D2S2330 are false positives.

Disequilibrium was also positive in the stratified data for two additional markers: D2S325 and D2S117. D2S325 demonstrated positive disequilibrium in male-only pairs with global $\chi^2_{2df} \equiv 6.89$ ($0.05>p>0.02$). This disequilibrium was accounted for by the significant ($0.05>p>0.02$) distorted transmission of two alleles (200bp allele [allele 1] as listed in the GDB and 190bp allele [allele 4] as listed in the GDB), one of which (the 190 bp allele) demonstrated excess transmission. D2S117 demonstrated positive disequilibrium in female-only pairs with global $\chi^2_{3df} = 8.05$ ($0.05>p>0.02$) and in male/hip pairs with global $\chi^2_{4df} \equiv 12.89$ ($0.02>p>0.01$). The positive disequilibrium in the female-only pairs was accounted for by the significant distorted transmission of three alleles, two of which demonstrated excess transmission ($0.05>p>0.01$) (192 bp allele [allele 3] as listed in the GDB and 202 bp allele [allele 6] as listed

in the GDB) whilst the third demonstrated reduced transmission (0.05>p>0.02) (208bp allele [allele 9] as listed in the GDB). The positive disequilibrium in the male/hips was accounted for by a single allele (208bp allele [allele 9] as listed in the GDB): observed transmission in male/hips of 85 compared to expected transmission of 74.63 (χ^2_{1df} = 8.15, 0.01>p>0.001). Although our male-only pairs did not demonstrate significant disequilibrium at D2S117 the transmission of alleles did approach significance in this category with a global χ^2_{3df}= 7.73 (0.1>p>0.05). In these male-only pairs allele 208bp did demonstrate distorted transmission with an observed transmission of 128 compared to an expected transmission of 117.20 (χ^2_{1df}= 5.95, 0.02>p>0.01). Intriguingly, it was the 208 bp allele that demonstrated reduced transmission in our female-only pairs with an observed transmission of 206 compared to an expected transmission of 218.4 (χ^2_{1df}=4.26, 0.05>p>0.02).

Methods

Affected sibling-pairs

Families were recruited which contained at least two siblings two had undergone one or more total hip (THR) and/or total knee replacements (TKR) for idiopathic OA. Clinically these patients are at the severe end of the OA spectrum with advanced radiological changes. The idiopathic diagnosis was supported by clinical, radiological, operative and histological findings: patients who had undergone joint replacement surgery secondary to other factors, such as intra-articular fracture or rheumatoid arthritis, were excluded. Families were ascertained at three centres within the United Kingdom: The Nuffield Orthopaedic Centre in Oxford, the Royal Orthopaedic Hospital in Birmingham and Musgrave Park Hospital in Belfast. Idiopathic OA is typically a late-onset disease and parents of affected siblings are rarely available. Of the 481 families used in

the study none contained a parent who was able to participate. We therefore collected siblings who had not undergone joint replacement to assist in the determination of identity-by-descent (IBD) allele transmittance. The 481 families were comprised of 1052 affected individuals plus an additional 302 unaffected siblings (Table 1). 59.3% of the affected individuals were female, 40.7% were male. For each individual, 25ml of venous blood was collected into EDTA tubes and DNA was extracted by conventional techniques.

Markers and Genotyping

Our initial screening panel of 272 microsatellite markers was essentially the panel used by Reed *et al*[8]. The additional microsatellite markers used to provide a denser coverage of chromosome 11 were obtained from the GDB or from the ABI Prism Linkage Mapping Set (Version 2). The markers were amplified using conventional conditions with either the forward or the reverse primer in a PCR pair fluorescently labelled. The amplification products were electrophoresed through 6% acrylamide using an Applied Biosystems 377 Automated DNA Sequencer[B]. Alleles were sized using Applied Biosystems Genescan version 2.0.2. and Genotyper version 1.1 software.

Linkage and Linkage disequilibrium analysis

Initially error checking procedures were employed for all families for each marker. After identification of straightforward mis-inheritances, more subtle transmission errors were detected using the PedCheck program[17]. The entire family data set was tested with Relative which is able to produce a probability calculation (based on 50 or more unlinked markers) that full sibships are in fact half sibs or even unrelated (due to unknown adoption of laboratory error). All 481 families used in the study successfully progressed through these checks. In addition markers were checked for

having excess homozygotes, based on their allele frequencies and heterozygosities. Markers shown to be unreliable were eliminated from the study. Finally the marker data were haplotyped for each chromosome using Simwalk2. This checks for areas on the chromosome where excessive recombination events may alert us to genotyping errors or mis-assignment of a marker position.

Non parametric linkage analysis was performed using the SIBPAIR module of the ANALYZE package[18]. This module is able to use data from siblings to determine identity-by-descent (IBD) allele transmittance. In this way it is appropriate to our study since we were unable to collect parents of our affected siblings. In the linkage analysis, siblings who had not undergone joint replacement were given a clinical status of unknown. The SIBPAIR module produces a singlepoint LOD score and its asymptotic P-value. Allele frequencies were calculated from the input data using either GAS or Downfreq (part of the ANALYZE package). Subsequent multipoint analyses was performed using ASPEX which calculates its own allele frequencies from the data set, using a maximum likelihood method, and employs marker information across the chromosome simultanenously[19]. ASPEX produces maximum LOD score (MLS) under an additive model. In addition it produces an exclusion map along the entire chromosome based on a fixed value for λs.

We tested for linkage disequilibrium by the transmission disequilibrium test (TDT) using the TRANSMIT software program[16]. Alleles with a frequency <0.1 were pooled. A global χ^2 statistic was computed for each microsatellite. If a microsatellite was significant a χ^2 statistic was then computed for its individual alleles.

Stratification

We stratified for sex, for joint replaced (hip or knee) and for sex combined with joint replaced.

For those families with more than two affected siblings and in which the siblings were not of the same sex, the affected sibling of opposite sex to a same pair was given an affected status of unknown in the linkage analysis. In this way we were stratifying for sex whilst not excluding siblings who could be used to assist in the determination of identity-by-descent (IBD) allele transmittance.

A hip-only pair were siblings who had each undergone joint replacement of the hip only (mono or bi-lateral) whilst a knee-only pair had undergone joint replacement of the knee only (mono or bi-lateral). If an affected pair was composed of one sibling who had undergone joint replacement of one joint type only (hip or knee) whilst their affected sibling had undergone joint replacement of the hip and knee then that pair were excluded. For an affected trio, if a pair of the siblings had both undergone joint replacement of the same joint type only (hip or knee) whilst the third sibling had undergone both hip and knee replacement, then the concordant pair were used in the stratification study whilst the third sibling was given an unaffected status in the linkage analysis. Again, we were attempting to maximise our determination of IBD allele transmittance.

Table 1(a)

Table 1 a) The families used in stages 1 and 2 together with the combined total figure. Also listed are the pairs concordant for different stratification criteria. b) A list of the individuals in the study.

a)

Families

	Stage 1	Stage 2	Total
Families	297	184	481
sibling pair	265	150	415
sibling trio	23	27	50
sibling quattro	7	5	12
other[a]	2	2	4
Female only	132	64	196
pair	121	59	180
trio	11	5	16
Male only	61	41	102
pair	57	36	93
trio	3	5	8
other[a]	1	0	1
Hip only	194	117	311
pair	170	96	266
trio	18	17	35
quattro	4	2	6
other[a]	2	2	4
Knee only	34	20	54
pair	33	19	52
trio	1	1	2
Female/hip only	85	47	132
pair	77	46	123
trio	8	1	9
Female/knee only	16	5	21
pair	16	5	21
Male/hip only	45	27	72
pair	41	25	66
trio	3	2	5
other[a]	1	0	1
Male/knee only	4	4	8
pair	4	4	8

Table 1(b)

b)

Individuals

	Stage 1	Stage 2	Total
Affected individuals	641	411	1052
Female	394	230	624
Male	247	181	428
Hip only	479	309	788
Knee only	121	77	198
Hip & Knee	41	25	66
Female/hip only	287	172	459
Female/knee only	77	42	119
Female/hip & knee	30	16	46
Males/hip only	192	137	329
Males/knee only	44	35	79
Males/hip & knee	11	9	20
Additional siblings[b]	211	91	302
Female	107	49	156
Male	104	42	146

[a]Other relative pairs such as cousins, uncles, aunts.
[b]Since our families lack parents, siblings who had not undergone joint replacement surgery were collected to assist in the determination of parental genotypes. These siblings were given an "unknown" clinical status in the linkage analysis.

Table 2 LOD scores and p-values for all markers that had a p≤0.05 in screen 1, for these markers in screen 2 and for screens 1 and 2 combined (* = p≤0.05).

	STAGE 1		STAGE 2		COMBINED	
Marker	p-value	LOD	p-value	LOD	p-value	LOD
D2S202	0.036*	0.70	0.07	0.49	0.009*	1.21
D3S1266	0.017*	0.96	0.5	0.00	0.082	0.42
D4S231	0.040*	0.67	0.5	0.00	0.33	0.04
D4S415	0.018*	0.95	0.33	0.04	0.025*	0.83
D6S260	0.050*	0.58	0.5	0.00	0.13	0.29
D6S273	0.016*	0.98	0.5	0.00	0.077	0.44
D6S286	0.030*	0.77	0.5	0.00	0.081	0.42
D6S281	0.046*	0.61	0.45	0.00	0.062	0.52
D7S669	0.018*	0.94	0.25	0.10	0.021*	0.90
D7S530	0.006*	1.33	0.41	0.01	0.013*	1.09
D11S907	0.050*	0.58	0.12	0.31	0.019*	0.92
D11S903	0.017*	0.97	0.07	0.49	0.004*	1.45
D11S901	0.0004*	2.45	0.10	0.37	0.0003*	2.51
D17S807	0.014*	1.03	0.5	0.00	0.15	0.24
D17S789	0.010*	1.16	0.5	0.00	0.071	0.47
DXS1068	0.020*	0.84	0.5	0.00	0.10	0.35

Table 3 LOD scores and p-value for stages 1 and 2 combined for the chromosome 2 markers.

Marker	cM from 2p telomere	p-value	LOD
D2S139	122	0.50	0.00
D2S160	145	0.041*	0.65
D2S114	167	0.050*	0.58
D2S142	186	0.28	0.07
D2S2330	194	0.19	0.16
D2S326	203	0.07	0.45
D2S364	212	0.08	0.41
D2S117	221	0.21	0.14
D2S202	223	0.009*	1.21
D2S72	225	0.007*	1.26
D2S325	231	0.09	0.39
D2S2382	242	0.31	0.05
D2S159	259	0.5	0.00

Table 4 Stratification of stages 1 and 2 combined of the chromosome 2 markers for affected female-only pairs, affected male-only pairs, hip-only pairs, knee-only pairs, affected females with hip-only disease and affected males with hip-only disease.

		FEMALES		MALES		HIPS		KNEES		FEMALE/HIPS		MALE/HIPS	
Marker	cM	p-value	LOD	p-value	LOD	p-value	LOD	p-value	LOD	p-value	LOD	p-value	LOD
D2S139	122	0.50	0.00	0.50	0.00	0.50	0.00	0.50	0.00	0.50	0.00	0.50	0.00
D2S160	145	0.32	0.04	0.35	0.03	0.028*	0.79	0.29	0.07	0.080	0.42	0.50	0.00
D2S114	167	0.29	0.06	0.11	0.32	0.022*	0.88	0.50	0.00	0.11	0.33	0.30	0.06
D2S142	186	0.50	0.00	0.052	0.57	0.099	0.36	0.50	0.00	0.50	0.00	0.008*	1.28
D2S2330	194	0.28	0.07	0.14	0.25	0.12	0.31	0.50	0.00	0.36	0.03	0.066	0.49
D2S326	203	0.019*	0.94	0.18	0.18	0.093	0.38	0.50	0.00	0.032*	0.74	0.13	0.28
D2S364	212	0.38	0.02	0.054	0.56	0.067	0.49	0.070	0.47	0.20	0.15	0.043*	0.64
D2S117	221	0.43	0.01	0.18	0.19	0.047*	0.61	0.14	0.25	0.19	0.16	0.037*	0.70
D2S202	223	0.40	0.01	0.0021*	1.79	0.0005*	2.36	0.34	0.03	0.20	0.15	0.0016*	1.89
D2S72	225	0.22	0.13	0.0025*	1.70	0.017*	0.97	0.22	0.13	0.11	0.34	0.0066*	1.34
D2S325	231	0.09	0.39	0.46	0.00	0.017*	0.98	0.50	0.00	0.13	0.28	0.30	0.06
D2S2382	242	0.49	0.00	0.47	0.00	0.26	0.09	0.17	0.20	0.16	0.22	0.50	0.00
D2S159	259	0.50	0.00	0.43	0.01	0.50	0.00	0.50	0.00	0.25	0.10	0.36	0.03

Table 5 Stratification of stages 1 and 2 combined of the 12 markers that had a p≤0.05 for stage 1 (excluding the chromosome 2 and 11 markers) for affected female-only pairs, affected male-only pairs, hip-only pairs, knee-only pairs, affected females with hip-only disease and affected males with hip-only disease.

	All Families		Females		Males		Hips		Knees		Female/Hips		Male/Hips	
Marker	p	LS	p	LS	p	LS	p	LS	p	LS	p	LS	p	LS
D3S1266	0.082	0.42	0.0060*	1.37	0.25	0.09	0.29	0.07	0.0035*	1.58	0.0067*	1.33	0.42	0.01
D4S231	0.33	0.04	0.045*	0.62	0.16	0.21	0.066	0.49	0.10	0.35	0.0052*	1.42	0.26	0.09
D4S415	0.025*	0.83	0.15	0.22	0.11	0.34	0.20	0.16	0.12	0.30	0.50	0.00	0.24	0.11
D6S260	0.13	0.29	0.16	0.22	0.50	0.00	0.19	0.17	0.50	0.00	0.50	0.00	0.33	0.04
D6S273	0.077	0.44	0.0034*	1.59	0.50	0.00	0.18	0.18	0.044*	0.63	0.0053*	1.42	0.5	0.00
D6S286	0.081	0.42	0.13	0.27	0.35	0.03	0.016*	0.99	0.50	0.00	0.058	0.54	0.36	0.03
D6S281	0.062	0.52	0.37	0.03	0.50	0.00	0.11	0.33	0.40	0.01	0.17	0.20	0.50	0.00
D7S669	0.021*	0.90	0.071	0.47	0.081	0.42	0.010*	1.16	0.39	0.02	0.032*	0.74	0.16	0.22
D7S530	0.013*	1.09	0.090	0.39	0.15	0.24	0.034*	0.72	0.17	0.19	0.18	0.18	0.29	0.07
D17S807	0.15	0.24	0.32	0.05	0.5	0.00	0.12	0.30	0.50	0.00	0.18	0.18	0.25	0.10
D17S789	0.071	0.47	0.12	0.29	0.39	0.02	0.25	0.10	0.021*	0.90	0.28	0.07	0.20	0.16
DXS1068	0.10	0.35	0.20	0.16	0.049*	0.60	0.19	0.17	0.26	0.09	0.50	0.00	0.02*	0.90

Table 6
TRANSMIT of chromosome 2.
$\chi 2$ analysis
(figures in brackets = degrees of freedom)

Locus	All data	Females-only	Males-only	Hips-only	Female/Hips	Male/Hips
D2S139	(5) 3.74	(5) 6.60	(5) 3.26	(5) 2.62	(5) 4.70	(5) 4.83
D2S160	(3) 3.19	(4) 1.80	(3) 3.51	(3) 5.69	(4) 5.96	(4) 3.32
D2S114	(4) 1.83	(5) 2.74	(4) 4.36	(4) 2.88	(5) 4.03	(4) 3.77
D2S142	(4) 2.52	(4) 2.18	(4) 3.52	(4) 0.87	(4) 1.25	(4) 2.44
D2S2330	(5) 11.43*	(5) 7.90	(5) 6.67	(5) 10.91	(5) 8.12	(4) 12.14*
D2S326	(4) 5.04	(5) 6.78	(4) 2.98	(4) 2.80	(4) 1.11	(4) 7.07
D2S364	(4) 5.59	(4) 3.89	(4) 3.43	(4) 4.37	(4) 1.28	(4) 4.42
D2S117	(3) 5.38	(3) 8.05*	(3) 7.73	(3) 5.31	(3) 5.19	(4) 12.89*
D2S202	(3) 3.59	(3) 3.98	(3) 0.50	(3) 3.01	(3) 6.32	(3) 1.45
D2S72	(3) 5.02	(4) 5.34	(3) 0.56	(3) 2.82	(4) 5.19	(3) 2.04
D2S325	(2) 3.51	(2) 1.00	(2) 6.89*	(2) 3.12	(3) 1.30	(3) 5.61
D2S2382	(2) 2.30	(2) 3.72	(2) 2.42	(2) 1.26	(2) 3.42	(2) 1.80
D25159	(4) 1.40	(4) 3.92	(4) 1.94	(4) 0.90	(4) 4.49	(4) 1.15